CLAIMS

Having thus described our invention, what we claim as new and desire to secure by Letters Patent is as follows:

1. A method of resource allocation to yield a benefit comprising the steps of:

generating an input time-customer matrix of demands for resources where a benefit function is known in advance; and

producing from the input matrix an output time-customer matrix of allocations of resources to customers to realize a benefit.

2. The method of resource allocation as recited in claim 1, wherein resource allocation is done to maximize a benefit.

3. The method of resource allocation as recited in claim 1, wherein the benefit is a tangible benefit.

4. The method of resource allocation as recited in claim 3, wherein the tangible benefit is a profit and resource allocation is done to maximize the profit.

5. The method of resource allocation as recited in claim 1, wherein the benefit is an intangible benefit.

6. The method of resource allocation as recited in claim 5, wherein the intangible benefit is customer satisfaction and resource allocation is done to maximize customer satisfaction.

7. The method of resource allocation as recited in claim 1, wherein the resource is computer cycles and resource allocation is done to more efficiently solve computationally intensive problems.

8. A method of resource allocation to yield a benefit comprising the steps of:

choosing a state s_t for each time t so as yield a benefit where all the state sets and the benefit function are known in advance;

reducing the problem to the analogous maximum-cost network flow problem by

constructing a directed network with s "rails", one per site, each rail being a chain of edges each representing one time step, flow along a rail representing an allocation of resources to a corresponding site,

constructing a set of "free pool" nodes, one per time step, through which flow will pass when resources are reallocated from one site to another,

for a demand matrix $d_{i,t}$, $1 \le i \le s$, $1 \le t \le T$, constructing nodes $n_{i,t}$, $1 \le i \le s$, $0 \le t \le T$, along with nodes f_t, $1 \le t \le T$, and for each site s and each time step t, constructing three edges from $n_{i,t-1}$ to $n_{i,t}$, wherein the first edge has capacity $\lfloor d_{i,t} \rfloor$ and cost $r_{i,t}$, the second edge has capacity one and cost $r_{i,t} \cdot (d_{i,t} - \lfloor d_{i,t} \rfloor)$, and the third edge has infinite capacity and cost zero, flow along the first edge representing a benefit of allocating resources s to site i during time step t, up to the integer part of $d_{i,t}$, flow along the second edge representing a remaining benefit, $r_{i,t}$, times a fractional part of $d_{i,t}$ to be collected by one more resource, and flow along the third edge representing that extra resources can remain allocated to s but do not collect any benefit,

constructing edges of infinite capacity and cost zero from $n_{i,t-1}$ to f_t

and from f_t to $n_{i,t}$, for each $1 \le t \le T$ and each $1 \le i \le s$ which represent a movement of servers from one site to another,

constructing a source into which a flow k is injected, with infinite capacity zero cost edges to each $n_{i,0}$, and a sink with infinite capacity zero cost edges from each $n_{i,T}$; and

solving the maximum-cost network flow problem and allocating resources.

9. The method of resource allocation as recited in claim 8, wherein resource allocation is done to maximize a benefit.

10. The method of resource allocation as recited in claim 8, wherein the benefit is a tangible benefit.

11. The method of resource allocation as recited in claim 10, wherein the tangible benefit is a profit and resource allocation is done to maximize the profit.

12. The method of resource allocation as recited in claim 8, wherein the benefit is an intangible benefit.

13. The method of resource allocation as recited in claim 12, wherein the intangible benefit is customer satisfaction and resource allocation is done to maximize customer satisfaction.

14. The method of resource allocation as recited in claim 8, wherein the resource is computer cycles and resource allocation is done to more efficiently solve computationally intensive problems.

15. A method for server allocation in a Web server "farm" based on information regarding future loads to achieve close to greatest possible revenue based on an assumption that revenue is proportional to the utilization of servers and differentiated by customer class comprising the steps of:

modeling the server allocation problem mathematically;

in the model, dividing time into intervals of fixed length based on the assumption that each site's demand is uniformly spread throughout each such interval;

maintaining server allocations fixed for the duration of an interval, servers being reallocated only at the beginning of an interval, and a reallocated server being unavailable for the length of the interval during which it is reallocated providing time to "scrub" the old site (customer data) to which the server was allocated, to reboot the server and to load the new site to which the server has been allocated, each server having a rate of requests it can serve in a time interval and customers share servers only in the sense of using the same servers at different times, but do not use the same servers at the same time;

associating each customer's demand with a benefit gained by the service provider in case a unit demand is satisfied and finding a time-varying server allocation that would maximized benefit gained by satisfying sites' demand; and

reducing to a minimum-cost network flow problem and solving in polynomial time.